Supporting information for:

Compressed Sensing for the Fast Computation of Matrices: Application to Molecular Vibrations

Jacob N. Sanders, Xavier Andrade, and Alán Aspuru-Guzik[*]

Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, United States

E-mail: aspuru@chemistry.harvard.edu

Further Comparison of Vibrational Frequencies

One reasonable question is whether the compressed sensing procedure outlined in the main article accurately recovers changes in the vibrational modes (i.e. the eigenvectors of the Hessian matrix) or whether it is helpful only in recovering changes in the vibrational frequencies (i.e. only the diagonal elements on the Hessian matrix). To address this question, we evaluated the true quantum mechanical (QM) vibrational frequencies along the vibrational modes obtained by the cheap molecular mechanics (MM) calculation and compared the results with compressed sensing. The results for anthracene are shown in Fig. S1. In particular, the figure compares the MM frequencies, the QM frequencies along the MM modes, and the compressed sensing frequencies with 35% sampling of the QM calculations according to the procedure described in the main article. (Note that the two plots present the same data with different frequency scales on the y-axis.)
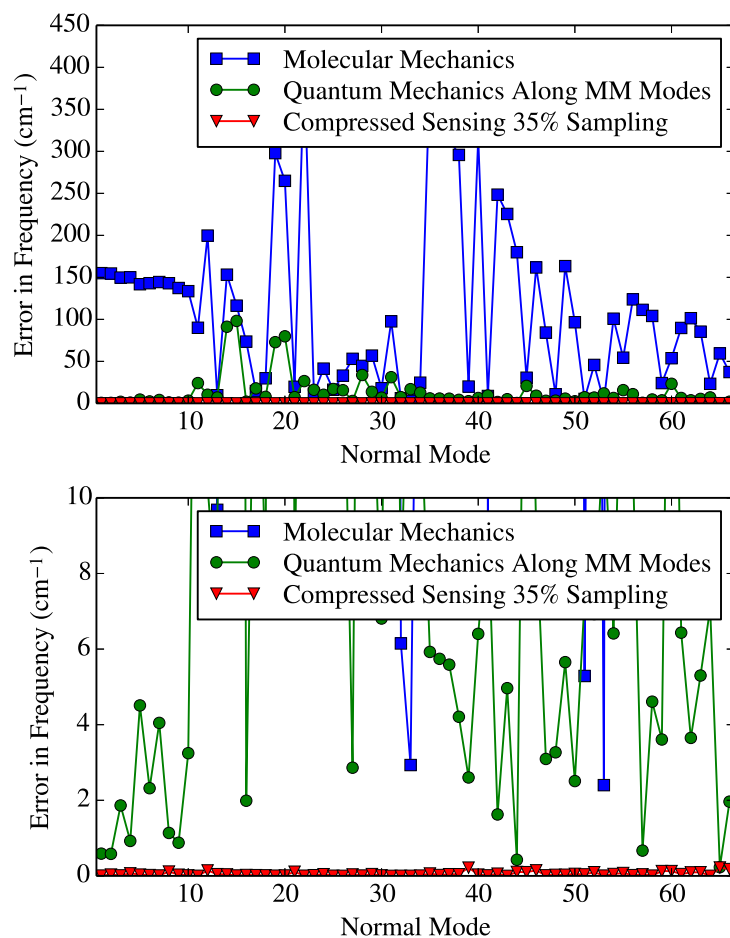
---

[*]To whom correspondence should be addressed

Figure S1: The figure compares the MM frequencies, the QM frequencies along the MM modes, and the compressed sensing frequencies with 35% sampling of the QM calculations according to the procedure described in the main article for anthracene. (Note that the two plots present the same data with different frequency scales on the y-axis.) The QM frequencies along the MM modes are much more accurate than the MM frequencies themselves, but only the compressed sensing calculation provides acceptable chemical accuracy.

The top plot shows that computing the QM frequencies along the MM modes provides a large improvement on the MM frequencies themselves, as expected since the eigenvalue calculation is quantum mechanical in nature and only the eigenvectors are approximate. However, substantial errors up to $100~\mathrm{cm}^{-1}$ remain in some of the frequencies so this approach does not suffice for chemical accuracy. In contrast, as the bottom plot shows, the compressed sensing frequencies are nearly exact, and are far more accurate than the QM frequencies computed along the MM modes. This result provides concrete numerical evidence that compressed sensing accurately recovers changes not only in the eigenvalues (i.e. diagonal elements of the Hessian matrix) but also in the vibrational modes themselves (i.e. the eigenvectors). As can be seen, only the compressed sensing approach, with accurate recovery of both vibrational modes and frequencies, suffices for chemical accuracy.

We also point out that as a method, computing the QM frequencies along the MM modes is based on an uncontrolled approximation, and the method can fail if the normal modes are predicted badly. With compressed sensing, a bad prediction of the normal modes is simply reflected in a more expensive calculation, but the results would still be accurate to the QM level.

Convergence Criterion

For practical implementation of the compressed sensing procedure (steps 1–7 in the main article), an important requirement is a convergence criterion; namely, how do you know when the procedure is done? In particular, without comparing to a full reference calculation, how do you know when enough columns of the Hessian have been sampled to achieve accurate vibrational modes and frequencies?

To address this question, we repeatedly applied the compressed sensing procedure (steps 1–7 in the main article) to the quantum mechanical Hessian matrix of anthracene, and each time we sampled one additional column. For each additional column sampled, we computed

the relative Frobenius norm (vectorial 2-norm) error against the full reference calculation,

$$\text{Overall Error}_n = \frac{||H_n - H_{\text{ref}}||_2}{||H_{\text{ref}}||_2}, \tag{1}$$

where H_n is the Hessian recovered via compressed sensing with n columns sampled in the measurement basis. We also computed the relative Frobenius norm error against the calculation in the previous step with one less sampled column,

$$\text{Successive Error}_n = \frac{||H_n - H_{n-1}||_2}{||H_n||_2}. \tag{2}$$

Fig. S2 shows both types of error plotted as a function of the number of sampled columns n. As the figure shows, the two errors track each other closely (never differing by more
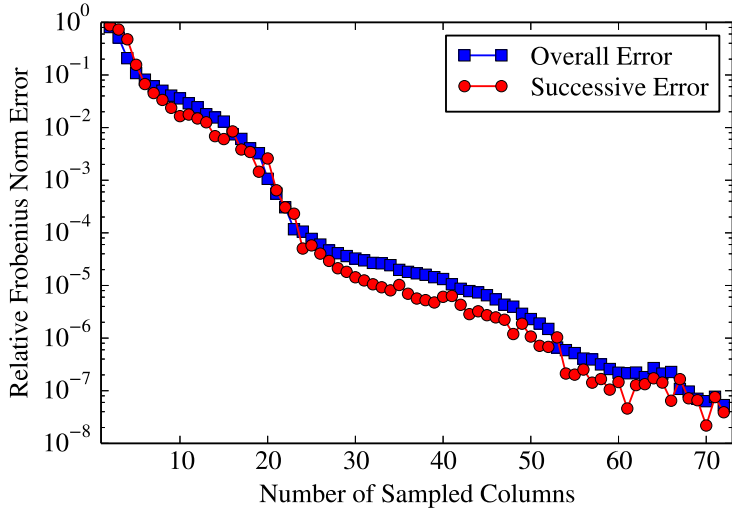

Figure S2: The figure shows both the overall error and the successive error in the Hessian of anthracene obtained via the compressed sensing procedure described in the main article as additional columns are sampled. Because the two errors track each other closely, the successive error can be used as a basis for developing a convergence criterion in the absence of a full reference calculation.

than an order of magnitude), but the successive error can be computed without access to a full reference calculation. Small successive errors indicate that the Hessian matrix is no longer changing when additional columns are sampled, and show that the calculation has

converged. Hence, just as in many other numerical methods, the successive error provides a measure for developing a convergence criterion. For example, to assure robustness, one might demand small successive errors over three or four consecutive steps before declaring the calculation complete.